

Exhibit 15

THIRD EDITION

MODERN EPIDEMIOLOGY

Kenneth J. Rothman

Vice President, Epidemiology Research
RTI Health Solutions
Professor of Epidemiology and Medicine
Boston University
Boston, Massachusetts

Sander Greenland

Professor of Epidemiology and Statistics
University of California
Los Angeles, California

Timothy L. Lash

Associate Professor of Epidemiology and Medicine
Boston University
Boston, Massachusetts



Wolters Kluwer | Lippincott Williams & Wilkins

Health

Philadelphia • Baltimore • New York • London
Buenos Aires • Hong Kong • Sydney • Tokyo

**Plaintiff's Exhibit
No.**

P-520

exhibits.law.com

Acquisitions Editor: Sonya Seigafuse
Developmental Editor: Louise Bierig
Project Manager: Kevin Johnson
Senior Manufacturing Manager: Ben Rivera
Marketing Manager: Kimberly Schonberger
Art Director: Risa Clow
Compositor: Aptara, Inc.

© 2008 by LIPPINCOTT WILLIAMS & WILKINS
530 Walnut Street
Philadelphia, PA 19106 USA
LWW.com

All rights reserved. This book is protected by copyright. No part of this book may be reproduced in any form or by any means, including photocopying, or utilized by any information storage and retrieval system without written permission from the copyright owner, except for brief quotations embodied in critical articles and reviews. Materials appearing in this book prepared by individuals as part of their official duties as U.S. government employees are not covered by the above-mentioned copyright.

Printed in China

Library of Congress Cataloging-in-Publication Data

Rothman, Kenneth J.

Modern epidemiology / Kenneth J. Rothman, Sander Greenland, and Timothy L. Lash. — 3rd ed.

p. ; cm.

2nd ed. edited by Kenneth J. Rothman and Sander Greenland.

Includes bibliographical references and index.

ISBN-13: 978-0-7817-5564-1

ISBN-10: 0-7817-5564-6

1. Epidemiology—Statistical methods. 2. Epidemiology—Research—Methodology. I. Greenland, Sander, 1951- II. Lash, Timothy L. III. Title.

[DNLM: 1. Epidemiology. 2. Epidemiologic Methods. WA 105 R846m 2008]

RA652.2.M3R67 2008

614.4—dc22

2007036316

Care has been taken to confirm the accuracy of the information presented and to describe generally accepted practices. However, the authors, editors, and publisher are not responsible for errors or omissions or for any consequences from application of the information in this book and make no warranty, expressed or implied, with respect to the currency, completeness, or accuracy of the contents of the publication. Application of this information in a particular situation remains the professional responsibility of the reader.

The publishers have made every effort to trace copyright holders for borrowed material. If they have inadvertently overlooked any, they will be pleased to make the necessary arrangements at the first opportunity.

To purchase additional copies of this book, call our customer service department at (800) 638-3030 or fax orders to 1-301-223-2400. Lippincott Williams & Wilkins customer service representatives are available from 8:30 am to 6:00 pm, EST, Monday through Friday, for telephone access. Visit Lippincott Williams & Wilkins on the Internet: <http://www.lww.com>.

10 9 8 7 6 5 4

Contents

Preface and Acknowledgments Contributors

vii

ix

1 Introduction

Kenneth J. Rothman, Sander Greenland, and Timothy L. Lash

1

SECTION I

Basic Concepts

2 Causation and Causal Inference

Kenneth J. Rothman, Sander Greenland, Charles Poole, and Timothy L. Lash

5

3 Measures of Occurrence

Sander Greenland and Kenneth J. Rothman

32

4 Measures of Effect and Measures of Association

Sander Greenland, Kenneth J. Rothman, and Timothy L. Lash

51

5 Concepts of Interaction

Sander Greenland, Timothy L. Lash, and Kenneth J. Rothman

71

SECTION II

Study Design and Conduct

6 Types of Epidemiologic Studies

Kenneth J. Rothman, Sander Greenland, and Timothy L. Lash

87

7 Cohort Studies

Kenneth J. Rothman and Sander Greenland

100

8 Case-Control Studies

Kenneth J. Rothman, Sander Greenland, and Timothy L. Lash

111

9 Validity in Epidemiologic Studies

Kenneth J. Rothman, Sander Greenland, and Timothy L. Lash

128

10 Precision and Statistics in Epidemiologic Studies

Kenneth J. Rothman, Sander Greenland, and Timothy L. Lash

148

11 Design Strategies to Improve Study Accuracy

Kenneth J. Rothman, Sander Greenland, and Timothy L. Lash

168

12 Causal Diagrams

M. Maria Glymour and Sander Greenland

183

iii

SECTION III

Data Analysis

13	Fundamentals of Epidemiologic Data Analysis	213
	<i>Sander Greenland and Kenneth J. Rothman</i>	
14	Introduction to Categorical Statistics	238
	<i>Sander Greenland and Kenneth J. Rothman</i>	
15	Introduction to Stratified Analysis	258
	<i>Sander Greenland and Kenneth J. Rothman</i>	
16	Applications of Stratified Analysis Methods	283
	<i>Sander Greenland</i>	
17	Analysis of Polytomous Exposures and Outcomes	303
	<i>Sander Greenland</i>	
18	Introduction to Bayesian Statistics	328
	<i>Sander Greenland</i>	
19	Bias Analysis	345
	<i>Sander Greenland and Timothy L. Lash</i>	
20	Introduction to Regression Models	381
	<i>Sander Greenland</i>	
21	Introduction to Regression Modeling	418
	<i>Sander Greenland</i>	

SECTION IV

Special Topics

22	Surveillance	459
	<i>James W. Buehler</i>	
23	Using Secondary Data	481
	<i>Jørn Olsen</i>	
24	Field Methods in Epidemiology	492
	<i>Patricia Hartge and Jack Cahill</i>	
25	Ecologic Studies	511
	<i>Hal Morgenstern</i>	
26	Social Epidemiology	532
	<i>Jay S. Kaufman</i>	
27	Infectious Disease Epidemiology	549
	<i>C. Robert Horsburgh, Jr., and Barbara E. Mahon</i>	
28	Genetic and Molecular Epidemiology	564
	<i>Muin J. Khoury, Robert Millikan, and Marta Gwinn</i>	
29	Nutritional Epidemiology	580
	<i>Walter C. Willett</i>	
30	Environmental Epidemiology	598
	<i>Irva Hertz-Picciotto</i>	
31	Methodologic Issues in Reproductive Epidemiology	620
	<i>Clarice R. Weinberg and Allen J. Wilcox</i>	

Contents

Contents

v

32 Clinical Epidemiology

641

*Noel S. Weiss***33 Meta-Analysis**

652

Sander Greenland and Keith O'Rourke

References

683

Index

733

213

238

258

283

303

328

345

381

418

459

481

492

511

532

549

564

580

598

620

Preface and Acknowledgments

This third edition of *Modern Epidemiology* arrives more than 20 years after the first edition, which was a much smaller single-authored volume that outlined the concepts and methods of a rapidly growing discipline. The second edition, published 12 years later, was a major transition, as the book grew along with the field. It saw the addition of a second author and an expansion of topics contributed by invited experts in a range of subdisciplines. Now, with the help of a third author, this new edition encompasses a comprehensive revision of the content and the introduction of new topics that 21st century epidemiologists will find essential.

This edition retains the basic organization of the second edition, with the book divided into four parts. Part I (Basic Concepts) now comprises five chapters rather than four, with the relocation of Chapter 5, "Concepts of Interaction," which was Chapter 18 in the second edition. The topic of interaction rightly belongs with Basic Concepts, although a reader aiming to accrue a working understanding of epidemiologic principles could defer reading it until after Part II, "Study Design and Conduct." We have added a new chapter on causal diagrams, which we debated putting into Part I, as it does involve basic issues in the conceptualization of relations between study variables. On the other hand, this material invokes concepts that seemed more closely linked to data analysis, and assumes knowledge of study design, so we have placed it at the beginning of Part III, "Data Analysis." Those with basic epidemiologic background could read Chapter 12 in tandem with Chapters 2 and 4 to get a thorough grounding in the concepts surrounding causal and non-causal relations among variables. Another important addition is a chapter in Part III titled, "Introduction to Bayesian Statistics," which we hope will stimulate epidemiologists to consider and apply Bayesian methods to epidemiologic settings. The former chapter on sensitivity analysis, now entitled "Bias Analysis," has been substantially revised and expanded to include probabilistic methods that have entered epidemiology from the fields of risk and policy analysis. The rigid application of frequentist statistical interpretations to data has plagued biomedical research (and many other sciences as well). We hope that the new chapters in Part III will assist in liberating epidemiologists from the shackles of frequentist statistics, and open them to more flexible, realistic, and deeper approaches to analysis and inference.

As before, Part IV comprises additional topics that are more specialized than those considered in the first three parts of the book. Although field methods still have wide application in epidemiologic research, there has been a surge in epidemiologic research based on existing data sources, such as registries and medical claims data. Thus, we have moved the chapter on field methods from Part II into Part IV, and we have added a chapter entitled, "Using Secondary Data." Another addition is a chapter on social epidemiology, and coverage on molecular epidemiology has been added to the chapter on genetic epidemiology. Many of these chapters may be of interest mainly to those who are focused on a particular area, such as reproductive epidemiology or infectious disease epidemiology, which have distinctive methodologic concerns, although the issues raised are well worth considering for any epidemiologist who wishes to master the field. Topics such as ecologic studies and meta-analysis retain a broad interest that cuts across subject matter subdisciplines. Screening had its own chapter in the second edition; its content has been incorporated into the revised chapter on clinical epidemiology.

The scope of epidemiology has become too great for a single text to cover it all in depth. In this book, we hope to acquaint those who wish to understand the concepts and methods of epidemiology with the issues that are central to the discipline, and to point the way to key references for further study. Although previous editions of the book have been used as a course text in many epidemiology

teaching programs, it is not written as a text for a specific course, nor does it contain exercises or review questions as many course texts do. Some readers may find it most valuable as a reference or supplementary-reading book for use alongside shorter textbooks such as Kelsey et al. (1996), Szklo and Nieto (2000), Savitz (2001), Koepsell and Weiss (2003), or Checkoway et al. (2004). Nonetheless, there are subsets of chapters that could form the textbook material for epidemiologic methods courses. For example, a course in epidemiologic theory and methods could be based on Chapters 1 through 12, with a more abbreviated course based on Chapters 1 through 4 and 6 through 11. A short course on the foundations of epidemiologic theory could be based on Chapters 1 through 5 and Chapter 12. Presuming a background in basic epidemiology, an introduction to epidemiologic data analysis could use Chapters 9, 10, and 12 through 19, while a more advanced course detailing causal and regression analysis could be based on Chapters 2 through 5, 9, 10, and 12 through 21. Many of the other chapters would also fit into such suggested chapter collections, depending on the program and the curriculum.

Many topics are discussed in various sections of the text because they pertain to more than one aspect of the science. To facilitate access to all relevant sections of the book that relate to a given topic, we have indexed the text thoroughly. We thus recommend that the index be consulted by those wishing to read our complete discussion of specific topics.

We hope that this new edition provides a resource for teachers, students, and practitioners of epidemiology. We have attempted to be as accurate as possible, but we recognize that any work of this scope will contain mistakes and omissions. We are grateful to readers of earlier editions who have brought such items to our attention. We intend to continue our past practice of posting such corrections on an internet page, as well as incorporating such corrections into subsequent printings. Please consult <<http://www.lww.com/ModernEpidemiology>> to find the latest information on errata.

We are also grateful to many colleagues who have reviewed sections of the current text and provided useful feedback. Although we cannot mention everyone who helped in that regard, we give special thanks to Onyebuchi Arah, Matthew Fox, Jamie Gradus, Jennifer Hill, Katherine Hoggatt, Marshal Joffe, Ari Lipsky, James Robins, Federico Soldani, Henrik Toft Sørensen, Soe Thwin and Tyler VanderWeele. An earlier version of Chapter 18 appeared in the *International Journal of Epidemiology* (2006;35:765–778), reproduced with permission of Oxford University Press. Finally, we thank Mary Anne Armstrong, Alan Dyer, Gary Friedman, Ulrik Gerdes, Paul Sorlie, and Katsuhiko Yano for providing unpublished information used in the examples of Chapter 33.

Kenneth J. Rothman
Sander Greenland
Timothy L. Lash

Preface and Acknowledgments

nor does it contain exercises or
it most valuable as a reference
s such as Kelsey et al. (1996),
, or Checkoway et al. (2004).
book material for epidemiologic
and methods could be based on
apters 1 through 4 and 6 through
be based on Chapters 1 through
n introduction to epidemiologic
more advanced course detailing
gh 5, 9, 10, and 12 through 21.
er collections, depending on the

se they pertain to more than one
of the book that relate to a given
that the index be consulted by

ers, students, and practitioners
ible, but we recognize that any
e grateful to readers of earlier
ad to continue our past practice
orporating such corrections into
nEpidemiology> to find the latest

sections of the current text and
e who helped in that regard, we
Gradus, Jennifer Hill, Katherine
ldani, Henrik Toft Sørensen, Soe
pter 18 appeared in the *Interna-*
with permission of Oxford Uni-
er, Gary Friedman, Ulrik Gerdes,
ormation used in the examples of

*Kenneth J. Rothman
Sander Greenland
Timothy L. Lash*

Contributors

James W. Buehler
Research Professor
Department of Epidemiology
Rollins School of Public Health
Emory University
Atlanta, Georgia

Jack Cahill
Vice President
Department of Health Studies Sector
Westat, Inc.
Rockville, Maryland

Sander Greenland
Professor of Epidemiology and
Statistics
University of California
Los Angeles, California

M. Maria Glymour
Robert Wood Johnson Foundation Health
and Society Scholar
Department of Epidemiology
Mailman School of Public Health
Columbia University
New York, New York
Department of Society, Human Development
and Health
Harvard School of Public Health
Boston, Massachusetts

Marta Gwinn
Associate Director
Department of Epidemiology
National Office of Public Health
Genomics
Centers for Disease Control and
Prevention
Atlanta, Georgia

Patricia Hartge
Deputy Director
Department of Epidemiology and
Biostatistics Program
Division of Cancer Epidemiology and Genetics
National Cancer Institute,
National Institutes of Health
Rockville, Maryland

Irva Hertz-Picciotto
Professor
Department of Public Health
University of California, Davis
Davis, California

C. Robert Horsburgh, Jr.
Professor of Epidemiology,
Biostatistics and Medicine
Department Epidemiology
Boston University School of Public Health
Boston, Massachusetts

Jay S. Kaufman
Associate Professor
Department of Epidemiology
University of North Carolina at Chapel Hill,
School of Public Health
Chapel Hill, North Carolina

Muin J. Khoury
Director
National Office of Public Health Genomics
Centers for Disease Control and Prevention
Atlanta, Georgia

Timothy L. Lash
Associate Professor of Epidemiology
and Medicine
Boston University
Boston, Massachusetts

x

Barbara E. Mahon

Assistant Professor
Department of Epidemiology and Pediatrics
Boston University
Novartis Vaccines and Diagnostics
Boston, Massachusetts

Robert C. Millikan

Professor
Department of Epidemiology
University of North Carolina at Chapel Hill,
School of Public Health
Chapel Hill, North Carolina

Hal Morgenstern

Professor and Chair
Department of Epidemiology
University of Michigan School of
Public Health
Ann Arbor, Michigan

Jørn Olsen

Professor and Chair
Department of Epidemiology
UCLA School of Public Health
Los Angeles, California

Keith O'Rourke

Visiting Assistant Professor
Department of Statistical Science
Duke University
Durham, North Carolina
Adjunct Professor
Department of Epidemiology and
Community Medicine
University of Ottawa
Ottawa, Ontario
Canada

Charles Poole

Associate Professor
Department of Epidemiology
University of North Carolina at Chapel Hill,
School of Public Health
Chapel Hill, North Carolina

Kenneth J. Rothman

Vice President, Epidemiology Research
RTI Health Solutions
Professor of Epidemiology and Medicine
Boston University
Boston, Massachusetts

Clarice R. Weinberg

National Institute of Environmental
Health Sciences
Biostatistics Branch
Research Triangle Park, North Carolina

Noel S. Weiss

Professor
Department of Epidemiology
University of Washington
Seattle, Washington

Allen J. Wilcox

Senior Investigator
Epidemiology Branch
National Institute of Environmental
Health Sciences/NIH
Durham, North Carolina

Walter C. Willett

Professor and Chair
Department of Nutrition
Harvard School of Public Health
Boston, Massachusetts

these hypotheses are not “proved” with the degree of absolute certainty that accompanies the proof of a mathematical theorem.

CAUSAL INFERENCE IN EPIDEMIOLOGY

Etiologic knowledge about epidemiologic hypotheses is often scant, making the hypotheses themselves at times little more than vague statements of causal association between exposure and disease, such as “smoking causes cardiovascular disease.” These vague hypotheses have only vague consequences that can be difficult to test. To cope with this vagueness, epidemiologists usually focus on testing the negation of the causal hypothesis, that is, the null hypothesis that the exposure does *not* have a causal relation to disease. Then, any observed association can potentially refute the hypothesis, subject to the assumption (auxiliary hypothesis) that biases and chance fluctuations are not solely responsible for the observation.

TESTS OF COMPETING EPIDEMIOLOGIC THEORIES

If the causal mechanism is stated specifically enough, epidemiologic observations can provide crucial tests of competing, non-null causal hypotheses. For example, when toxic-shock syndrome was first studied, there were two competing hypotheses about the causal agent. Under one hypothesis, it was a chemical in the tampon, so that women using tampons were exposed to the agent directly from the tampon. Under the other hypothesis, the tampon acted as a culture medium for staphylococci that produced a toxin. Both hypotheses explained the relation of toxic-shock occurrence to tampon use. The two hypotheses, however, led to opposite predictions about the relation between the frequency of changing tampons and the rate of toxic shock. Under the hypothesis of a chemical agent, more frequent changing of the tampon would lead to more exposure to the agent and possible absorption of a greater overall dose. This hypothesis predicted that women who changed tampons more frequently would have a higher rate than women who changed tampons infrequently. The culture-medium hypothesis predicts that women who change tampons frequently would have a lower rate than those who change tampons less frequently, because a short duration of use for each tampon would prevent the staphylococci from multiplying enough to produce a damaging dose of toxin. Thus, epidemiologic research, by showing that infrequent changing of tampons was associated with a higher rate of toxic shock, refuted the chemical theory in the form presented. There was, however, a third hypothesis that a chemical in some tampons (e.g., oxygen content) improved their performance as culture media. This chemical-promotor hypothesis made the same prediction about the association with frequency of changing tampons as the microbial toxin hypothesis (Lanes and Rothman, 1990).

Another example of a theory that can be easily tested by epidemiologic data relates to the observation that women who took replacement estrogen therapy had a considerably elevated rate of endometrial cancer. Horwitz and Feinstein (1978) conjectured a competing theory to explain the association: They proposed that women taking estrogen experienced symptoms such as bleeding that induced them to consult a physician. The resulting diagnostic workup led to the detection of endometrial cancer at an earlier stage in these women, as compared with women who were not taking estrogens. Horwitz and Feinstein argued that the association arose from this detection bias, claiming that without the bleeding-induced workup, many of these cancers would not have been detected at all. Many epidemiologic observations were used to evaluate these competing hypotheses. The detection-bias theory predicted that women who had used estrogens for only a short time would have the greatest elevation in their rate, as the symptoms related to estrogen use that led to the medical consultation tended to appear soon after use began. Because the association of recent estrogen use and endometrial cancer was the same in both long- and short-term estrogen users, the detection-bias theory was refuted as an explanation for all but a small fraction of endometrial cancer cases occurring after estrogen use. Refutation of the detection-bias theory also depended on many other observations. Especially important was the theory’s implication that there must be a huge reservoir of undetected endometrial cancer in the typical population of women to account for the much greater rate observed in estrogen users, an implication that was not borne out by further observations (Hutchison and Rothman, 1978).

The endometrial cancer example illustrates a critical point in understanding the process of causal inference in epidemiologic studies: Many of the hypotheses being evaluated in the interpretation of epidemiologic studies are auxiliary hypotheses in the sense that they are independent of the presence, absence, or direction of any causal connection between the study exposure and the disease. For example, explanations of how specific types of bias could have distorted an association between exposure and disease are the usual alternatives to the primary study hypothesis. Much of the interpretation of epidemiologic studies amounts to the testing of such auxiliary explanations for observed associations.

CAUSAL CRITERIA

In practice, how do epidemiologists separate causal from noncausal explanations? Despite philosophical criticisms of inductive inference, inductively oriented considerations are often used as criteria for making such inferences (Weed and Gorelic, 1996). If a set of necessary and sufficient causal criteria could be used to distinguish causal from noncausal relations in epidemiologic studies, the job of the scientist would be eased considerably. With such criteria, all the concerns about the logic or lack thereof in causal inference could be subsumed: It would only be necessary to consult the checklist of criteria to see if a relation were causal. We know from the philosophy reviewed earlier that a set of sufficient criteria does not exist. Nevertheless, lists of causal criteria have become popular, possibly because they seem to provide a road map through complicated territory, and perhaps because they suggest hypotheses to be evaluated in a given problem.

A commonly used set of criteria was based on a list of considerations or "viewpoints" proposed by Sir Austin Bradford Hill (1965). Hill's list was an expansion of a list offered previously in the landmark U.S. Surgeon General's report *Smoking and Health* (1964), which in turn was anticipated by the inductive canons of John Stuart Mill (1862) and the rules given by Hume (1739). Subsequently, others, especially Susser, have further developed causal considerations (Kaufman and Poole, 2000).

Hill suggested that the following considerations in attempting to distinguish causal from noncausal associations that were already "perfectly clear-cut and beyond what we would care to attribute to the play of chance": (1) strength, (2) consistency, (3) specificity, (4) temporality, (5) biologic gradient, (6) plausibility, (7) coherence, (8) experimental evidence, and (9) analogy. Hill emphasized that causal inferences cannot be based on a set of rules, condemned emphasis on statistical significance testing, and recognized the importance of many other factors in decision making (Phillips and Goodman, 2004). Nonetheless, the misguided but popular view that his considerations should be used as criteria for causal inference makes it necessary to examine them in detail.

Strength

Hill argued that strong associations are particularly compelling because, for weaker associations, it is "easier" to imagine what today we would call an unmeasured confounder that might be responsible for the association. Several years earlier, Cornfield et al. (1959) drew similar conclusions. They concentrated on a single hypothetical confounder that, by itself, would explain entirely an observed association. They expressed a strong preference for ratio measures of strength, as opposed to difference measures, and focused on how the observed estimate of a risk ratio provides a minimum for the association that a completely explanatory confounder must have with the exposure (rather than a minimum for the confounder-disease association). Of special importance, Cornfield et al. acknowledged that having only a weak association does not rule out a causal connection (Rothman and Poole, 1988). Today, some associations, such as those between smoking and cardiovascular disease or between environmental tobacco smoke and lung cancer, are accepted by most as causal even though the associations are considered weak.

Counterexamples of strong but noncausal associations are also not hard to find; any study with strong confounding illustrates the phenomenon. For example, consider the strong relation between Down syndrome and birth rank, which is confounded by the relation between Down syndrome and maternal age. Of course, once the confounding factor is identified, the association is diminished by controlling for the factor.

These examples remind us that a strong association is neither necessary nor sufficient for causality, and that weakness is neither necessary nor sufficient for absence of causality. A strong association

bears only on hypotheses that the association is entirely or partially due to unmeasured confounders or other source of modest bias.

Consistency

To most observers, consistency refers to the repeated observation of an association in different populations under different circumstances. Lack of consistency, however, does not rule out a causal association, because some effects are produced by their causes only under unusual circumstances. More precisely, the effect of a causal agent cannot occur unless the complementary component causes act or have already acted to complete a sufficient cause. These conditions will not always be met. Thus, transfusions can cause infection with the human immunodeficiency virus, but they do not always do so: The virus must also be present. Tampon use can cause toxic-shock syndrome, but only rarely, when certain other, perhaps unknown, conditions are met. Consistency is apparent only after all the relevant details of a causal mechanism are understood, which is to say very seldom. Furthermore, even studies of exactly the same phenomena can be expected to yield different results simply because they differ in their methods and random errors. Consistency serves only to rule out hypotheses that the association is attributable to some factor that varies across studies.

One mistake in implementing the consistency criterion is so common that it deserves special mention. It is sometimes claimed that a literature or set of results is inconsistent simply because some results are “statistically significant” and some are not. This sort of evaluation is completely fallacious even if one accepts the use of significance testing methods. The results (effect estimates) from a set of studies could all be identical even if many were significant and many were not, the difference in significance arising solely because of differences in the standard errors or sizes of the studies. Conversely, the results could be significantly in conflict even if all were all were nonsignificant individually, simply because in aggregate an effect could be apparent in some subgroups but not others (see Chapter 33). The fallacy of judging consistency by comparing *P*-values or statistical significance is not eliminated by “standardizing” estimates (i.e., dividing them by the standard deviation of the outcome, multiplying them by the standard deviation of the exposure, or both); in fact it is worsened, as such standardization can create differences where none exists, or mask true differences (Greenland et al., 1986, 1991; see Chapters 21 and 33).

Specificity

The criterion of specificity has two variants. One is that a cause leads to a single effect, not multiple effects. The other is that an effect has one cause, not multiple causes. Hill mentioned both of them. The former criterion, specificity of effects, was used as an argument in favor of a causal interpretation of the association between smoking and lung cancer and, in an act of circular reasoning, in favor of ratio comparisons and not differences as the appropriate measures of strength. When ratio measures were examined, the association of smoking to diseases looked “quantitatively specific” to lung cancer. When difference measures were examined, the association appeared to be nonspecific, with several diseases (other cancers, coronary heart disease, etc.) being at least as strongly associated with smoking as lung cancer was. Today we know that smoking affects the risk of many diseases and that the difference comparisons were accurately portraying this lack of specificity. Unfortunately, however, the historical episode of the debate over smoking and health is often cited today as justification for the specificity criterion and for using ratio comparisons to measure strength of association. The proper lessons to learn from that episode should be just the opposite.

Weiss (2002) argued that specificity can be used to distinguish some causal hypotheses from noncausal hypotheses, when the causal hypothesis predicts a relation with one outcome but no relation with another outcome. His argument is persuasive when, in addition to the causal hypothesis, one has an alternative noncausal hypothesis that predicts a nonspecific association. Weiss offered the example of screening sigmoidoscopy, which was associated in case-control studies with a 50% to 70% reduction in mortality from distal tumors of the rectum and tumors of the distal colon, within the reach of the sigmoidoscope, but no reduction in mortality from tumors elsewhere in the colon. If the effect of screening sigmoidoscopy were not specific to the distal colon tumors, it would lend support not to all noncausal theories to explain the association, as Weiss suggested, but only to those noncausal theories that would have predicted a nonspecific association. Thus, specificity can

come into play when it can be logically deduced from the causal hypothesis in question and when nonspecificity can be logically deduced from one or more noncausal hypotheses.

Temporality

Temporality refers to the necessity that the cause precede the effect in time. This criterion is inarguable, insofar as any claimed observation of causation must involve the putative cause C preceding the putative effect D. It does *not*, however, follow that a reverse time order is evidence against the hypothesis that C can cause D. Rather, observations in which C followed D merely show that C could not have caused D in these instances; they provide no evidence for or against the hypothesis that C can cause D in those instances in which it precedes D. Only if it is found that C cannot precede D can we dispense with the causal hypothesis that C *could* cause D.

Biologic Gradient

Biologic gradient refers to the presence of a dose-response or exposure-response curve with an expected shape. Although Hill referred to a "linear" gradient, without specifying the scale, a linear gradient on one scale, such as the risk, can be distinctly nonlinear on another scale, such as the log risk, the odds, or the log odds. We might relax the expectation from linear to strictly monotonic (steadily increasing or decreasing) or even further merely to monotonic (a gradient that never changes direction). For example, more smoking means more carcinogen exposure and more tissue damage, hence more opportunity for carcinogenesis. Some causal associations, however, show a rapid increase in response (an approximate threshold effect) rather than a strictly monotonic trend. An example is the association between DES and adenocarcinoma of the vagina. A possible explanation is that the doses of DES that were administered were all sufficiently great to produce the maximum effect from DES. Under this hypothesis, for all those exposed to DES, the development of disease would depend entirely on other component causes.

The somewhat controversial topic of alcohol consumption and mortality is another example. Death rates are higher among nondrinkers than among moderate drinkers, but they ascend to the highest levels for heavy drinkers. There is considerable debate about which parts of the J-shaped dose-response curve are causally related to alcohol consumption and which parts are noncausal artifacts stemming from confounding or other biases. Some studies appear to find only an increasing relation between alcohol consumption and mortality, possibly because the categories of alcohol consumption are too broad to distinguish different rates among moderate drinkers and nondrinkers, or possibly because they have less confounding at the lower end of the consumption scale.

Associations that do show a monotonic trend in disease frequency with increasing levels of exposure are not necessarily causal. Confounding can result in a monotonic relation between a noncausal risk factor and disease if the confounding factor itself demonstrates a biologic gradient in its relation with disease. The relation between birth rank and Down syndrome mentioned earlier shows a strong biologic gradient that merely reflects the progressive relation between maternal age and occurrence of Down syndrome.

These issues imply that the existence of a monotonic association is neither necessary nor sufficient for a causal relation. A nonmonotonic relation only refutes those causal hypotheses specific enough to predict a monotonic dose-response curve.

Plausibility

Plausibility refers to the scientific plausibility of an association. More than any other criterion, this one shows how narrowly systems of causal criteria are focused on epidemiology. The starting point is an epidemiologic association. In asking whether it is causal or not, one of the considerations we take into account is its plausibility. From a less parochial perspective, the entire enterprise of causal inference would be viewed as the act of determining how plausible a causal *hypothesis* is. One of the considerations we would take into account would be epidemiologic associations, if they are available. Often they are not, but causal inference must be done nevertheless, with inputs from toxicology, pharmacology, basic biology, and other sciences.

Just as epidemiology is not essential for causal inference, plausibility can change with the times. Sartwell (1960) emphasized this point, citing remarks of Cheever in 1861, who had been commenting on the etiology of typhus before its mode of transmission (via body lice) was known:

It could be no more ridiculous for the stranger who passed the night in the steerage of an emigrant ship to ascribe the typhus, which he there contracted, to the vermin with which bodies of the sick might be infested. An adequate cause, one reasonable in itself, must correct the coincidences of simple experience.

What was to Cheever an implausible explanation turned out to be the correct explanation, because it was indeed the vermin that caused the typhus infection. Such is the problem with plausibility: It is too often based not on logic or data, but only on prior beliefs. This is not to say that biologic knowledge should be discounted when a new hypothesis is being evaluated, but only to point out the difficulty in applying that knowledge.

The Bayesian approach to inference attempts to deal with this problem by requiring that one quantify, on a probability (0 to 1) scale, the certainty that one has in prior beliefs, as well as in new hypotheses. This quantification displays the dogmatism or open-mindedness of the analyst in a public fashion, with certainty values near 1 or 0 betraying a strong commitment of the analyst for or against a hypothesis. It can also provide a means of testing those quantified beliefs against new evidence (Howson and Urbach, 1993). Nevertheless, no approach can transform plausibility into an objective causal criterion.

Coherence

Taken from the U.S. Surgeon General's *Smoking and Health* (1964), the term *coherence* implies that a cause-and-effect interpretation for an association does not conflict with what is known of the natural history and biology of the disease. The examples Hill gave for coherence, such as the histopathologic effect of smoking on bronchial epithelium (in reference to the association between smoking and lung cancer) or the difference in lung cancer incidence by sex, could reasonably be considered examples of plausibility, as well as coherence; the distinction appears to be a fine one. Hill emphasized that the absence of coherent information, as distinguished, apparently, from the presence of conflicting information, should not be taken as evidence against an association being considered causal. On the other hand, the presence of conflicting information may indeed refute a hypothesis, but one must always remember that the conflicting information may be mistaken or misinterpreted. An example mentioned earlier is the "inhalation anomaly" in smoking and lung cancer, the fact that the excess of lung cancers seen among smokers seemed to be concentrated at sites in the upper airways of the lung. Several observers interpreted this anomaly as evidence that cigarettes were not responsible for the excess. Other observations, however, suggested that cigarette-borne carcinogens were deposited preferentially where the excess was observed, and so the anomaly was in fact consistent with a causal role for cigarettes (Wald, 1985).

Experimental Evidence

To different observers, experimental evidence can refer to clinical trials, to laboratory experiments with rodents or other nonhuman organisms, or to both. Evidence from human experiments, however, is seldom available for epidemiologic research questions, and animal evidence relates to different species and usually to levels of exposure very different from those that humans experience. Uncertainty in extrapolations from animals to humans often dominates the uncertainty of quantitative risk assessments (Freedman and Zeisel, 1988; Crouch et al., 1997).

To Hill, however, experimental evidence meant something else: the "experimental, or semi-experimental evidence" obtained from reducing or eliminating a putatively harmful exposure and seeing if the frequency of disease subsequently declines. He called this the strongest possible evidence of causality that can be obtained. It can be faulty, however, as the "semi-experimental" approach is nothing more than a "before-and-after" time trend analysis, which can be confounded or otherwise biased by a host of concomitant secular changes. Moreover, even if the removal of exposure does causally reduce the frequency of disease, it might not be for the etiologic reason hypothesized. The draining of a swamp near a city, for instance, would predictably and causally reduce the rate of yellow fever or malaria in that city the following summer. But it would be a mistake to call this observation the strongest possible evidence of a causal role of miasmas (Poole, 1999).

Analogy

Whatever insight might be derived from analogy is handicapped by the inventive imagination of scientists who can find analogies everywhere. At best, analogy provides a source of more elaborate hypotheses about the associations under study; absence of such analogies reflects only lack of imagination or experience, not falsity of the hypothesis.

We might find naive Hill's examples in which reasoning by analogy from the thalidomide and rubella tragedies made it more likely to him that other medicines and infections might cause other birth defects. But such reasoning is common; we suspect most people find it more credible that smoking might cause, say, stomach cancer, because of its associations, some widely accepted as causal, with cancers in other internal and gastrointestinal organs. Here we see how the analogy criterion can be at odds with either of the two specificity criteria. The more apt the analogy, the less specific are the effects of a cause or the less specific the causes of an effect.

Summary

As is evident, the standards of epidemiologic evidence offered by Hill are saddled with reservations and exceptions. Hill himself was ambivalent about their utility. He did not use the word *criteria* in the speech. He called them "viewpoints" or "perspectives." On the one hand, he asked, "In what circumstances can we pass from this observed *association* to a verdict of *causation*?" (emphasis in original). Yet, despite speaking of verdicts on causation, he disagreed that any "hard-and-fast rules of evidence" existed by which to judge causation: "None of my nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis and none can be required as a *sine qua non*" (Hill, 1965).

Actually, as noted above, the fourth viewpoint, temporality, is a *sine qua non* for causal explanations of observed associations. Nonetheless, it does not bear on the hypothesis that an exposure is capable of causing a disease in situations as yet unobserved (whether in the past or the future). For suppose every exposed case of disease ever reported had received the exposure after developing the disease. This reversed temporal relation would imply that exposure had not caused disease among these reported cases, and thus would refute the hypothesis that it had. Nonetheless, it would *not* refute the hypothesis that the exposure is *capable* of causing the disease, or that it had caused the disease in unobserved cases. It would mean only that we have no worthwhile epidemiologic evidence relevant to that hypothesis, for we had not yet seen what became of those exposed before disease occurred relative to those unexposed. Furthermore, what appears to be a causal sequence could represent reverse causation if preclinical symptoms of the disease lead to exposure, and then overt disease follows, as when patients in pain take analgesics, which may be the result of disease that is later diagnosed, rather than a cause.

Other than temporality, there is no necessary or sufficient criterion for determining whether an observed association is causal. Only when a causal hypothesis is elaborated to the extent that one can predict from it a particular form of consistency, specificity, biologic gradient, and so forth, can "causal criteria" come into play in evaluating causal hypotheses, and even then they do not come into play in evaluating the general hypothesis *per se*, but only some specific causal hypotheses, leaving others untested.

This conclusion accords with the views of Hume and many others that causal inferences cannot attain the certainty of logical deductions. Although some scientists continue to develop causal considerations as aids to inference (Susser, 1991), others argue that it is detrimental to cloud the inferential process by considering checklist criteria (Lanes and Poole, 1984). An intermediate, refutationist approach seeks to transform proposed criteria into deductive tests of causal hypotheses (Maclure, 1985; Weed, 1986). Such an approach helps avoid the temptation to use causal criteria simply to buttress pet theories at hand, and instead allows epidemiologists to focus on evaluating competing causal theories using crucial observations. Although this refutationist approach to causal inference may seem at odds with the common implementation of Hill's viewpoints, it actually seeks to answer the fundamental question posed by Hill, and the ultimate purpose of the viewpoints he promulgated:

What [the nine viewpoints] can do, with greater or less strength, is to help us to make up our minds on the fundamental question—is there any other way of explaining the set of facts before us, is there any other answer equally, or more, likely than cause and effect? (Hill, 1965)

The crucial phrase “equally or more likely than cause and effect” suggests to us a subjective assessment of the certainty, or probability of the causal hypothesis at issue relative to another hypothesis. Although Hill wrote at a time when expressing uncertainty as a probability was unpopular in statistics, it appears from his statement that, for him, causal inference is a subjective matter of degree of personal belief, certainty, or conviction. In any event, this view is precisely that of subjective Bayesian statistics (Chapter 18).

It is unsurprising that case studies (e.g., Weed and Gorelick, 1996) and surveys of epidemiologists (Holman et al., 2001) show, contrary to the rhetoric that often attends invocations of causal criteria, that epidemiologists have *not* agreed on a set of causal criteria or on how to apply them. In one study in which epidemiologists were asked to employ causal criteria to fictional summaries of epidemiologic literatures, the agreement was only slightly greater than would have been expected by chance (Holman et al., 2001). The typical use of causal criteria is to make a case for a position for or against causality that has been arrived at by other, unstated means. Authors pick and choose among the criteria they deploy, and define and weight them in *ad hoc* ways that depend only on the exigencies of the discussion at hand. In this sense, causal criteria appear to function less like standards or principles and more like values (Poole, 2001b), which vary across individual scientists and even vary within the work of a single scientist, depending on the context and time. Thus universal and objective causal criteria, if they exist, have yet to be identified.